

heritability,” or regard the higher estimates from SumHer as being closer to the “true value.” However, from the standpoint of estimands, such differences do not necessarily reflect the superiority or inferiority of methods; rather, they are more likely to arise because the definitions of heritability targeted by these methods are themselves not equivalent. In other words, these so-called “contradictions” largely stem from the incommensurability of the quantities being compared, rather than simple differences in estimation accuracy. Therefore, when interpreting results from different methods, priority should be given to identifying the estimand each method corresponds to, rather than making direct judgments based solely on numerical comparisons.

#### **4.4 Methodological implications and best practices**

From the perspective of method selection, GREML-based approaches tend to exhibit relatively stable performance under certain conditions. In particular, when the sample size is large (e.g.,  $N > 100,000$ ), the analysis focuses primarily on common variants ( $MAF > 0.01$ ), and genetic effects are approximately uniformly distributed, GREML and its approximations generally provide estimates with lower variance and greater robustness. This pattern has been empirically supported in large-scale datasets such as the UK Biobank (Hou et al., 2019). Under such conditions, the alignment between model assumptions and data characteristics is relatively strong, thereby reducing the risk of systematic bias.

However, when the genetic architecture deviates from these idealized conditions, GREML estimates may exhibit systematic underestimation. For example, when a trait is predominantly influenced by low-frequency or rare variants, when the LD structure is highly heterogeneous, or when genotyping data provide incomplete coverage of the underlying causal variation, the conventional GREML framework may fail to adequately capture these complexities. In such cases, extensions that incorporate LD and MAF stratification (e.g., GREML-LDMS), or methods that apply weighting schemes to effect sizes such as LDAK, can partially correct these biases and improve the interpretability of the estimates (Speed et al., 2017).

Based on these considerations, a single method is often insufficient to fully characterize the heritability structure of complex traits. A more robust strategy is therefore to adopt a multi-method analytical framework. In practice, GREML results may serve as a baseline estimate, while LDSC can be used for external validation based on summary statistics, and SumHer can be incorporated to assess sensitivity to assumptions about genetic architecture. Building upon this, further analyses may include LD-stratified approaches (e.g., GREML-LDMS) and the exclusion or separate evaluation of specific genomic regions (such as the MHC region), with consistency checks across methods used to identify potential sources of bias. Such an integrated strategy helps establish clearer correspondences among different estimands, reduces overreliance on any single method, and enhances the overall robustness and interpretability of inference.

#### **4.5 Broader implications for statistical genetics**

A key theoretical advancement of this study lies in reconceptualizing SNP heritability. Rather than treating it as a fixed and directly comparable single parameter, we define it as an estimand that depends on the specification of the statistical model. This shift in perspective is not only methodologically significant but also provides a new interpretative pathway for several long-standing debates in statistical genetics. Taking the “missing heritability” problem as an example, previous studies have often attributed discrepancies between different methods to unobserved genetic variation or limitations in sample size. However, to a considerable extent, these discrepancies arise because different models correspond to different estimands.

From this standpoint, the seemingly inconsistent results produced by different estimation methods can be reinterpreted as differences in estimation targets rather than estimation errors. This insight provides a theoretical foundation for integrating diverse statistical tools, allowing previously fragmented analytical frameworks to be understood within a unified conceptual system. At the same time, it offers clearer guidance for future methodological development: model design should not focus solely on improving estimation accuracy, but must also explicitly define the corresponding statistical object and its biological interpretation.